550970

(12) DEMANDE INTERNATIONALE PUBLIÉE EN VERTU DU TRAITÉ DE COOPÉRATION EN MATIÈRE DE BREVETS (PCT)

(19) Organisation Mondiale de la Propriété Intellectuelle
Bureau international

(43) Date de la publication internationale
14 octobre 2004 (14.10.2004)

PCT

(10) Numéro de publication internationale
WO 2004/088636 A1

(51) Classification internationale des brevets[7] : G10L 15/28

(21) Numéro de la demande internationale :
PCT/FR2004/000546

(22) Date de dépôt international : 8 mars 2004 (08.03.2004)

(25) Langue de dépôt : français

(26) Langue de publication : français

(30) Données relatives à la priorité :
03/03615 25 mars 2003 (25.03.2003) FR

(71) Déposant *(pour tous les États désignés sauf US)* : **FRANCE TELECOM** [FR/FR]; 6, place d'Alleray, F-75015 Paris (FR).

(72) Inventeurs; et

(75) Inventeurs/Déposants *(pour US seulement)* : **MONNE, Jean** [FR/FR]; 53, rue du Pré de Saint-Maur, F-22700 Perros Guirec (FR). **PETIT, Jean-Pierre** [FR/FR]; 10, lot Zant Erwan, F-22220 Minihy Treguier (FR). **BRISARD, Patrick** [FR/FR]; 5, allée du Cédre, F-92320 Chatillon (FR).

(74) Mandataires : **LOISEL, Bertrand** etc.; Cabinet Plasseraud, 65/67, rue de la Victoire, F-75440 Paris Cedex 09 (FR).

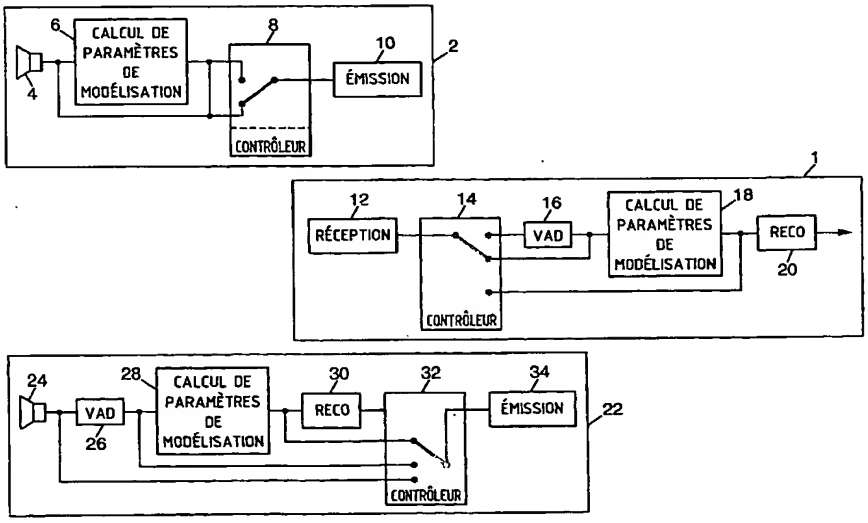
(81) États désignés *(sauf indication contraire, pour tout titre de protection nationale disponible)* : AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NA, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) États désignés *(sauf indication contraire, pour tout titre de protection régionale disponible)* : ARIPO (BW, GH, GM,

[Suite sur la page suivante]

(54) Title: DISTRIBUTED SPEECH RECOGNITION SYSTEM

(54) Titre : SYSTEME DE RECONNAISSANCE DE PAROLE DISTRIBUEE



6 / 18 / 28...CALCULATION OF MODELLING PARAMETERS
8 / 14 / 32...CONTROLLER
10 / 34...TRANSMISSION
12...RECEPTION

(57) Abstract: The invention relates to a distributed speech recognition system. The inventive system consists of: at least one user terminal comprising means for obtaining an audio signal to be recognised, parameter calculation means and control means which are used to select a signal to be transmitted; and a server comprising means for receiving the signal, parameter calculation means, recognition means and control means which are used to control the calculation means and the recognition means according to the signal received.

[Suite sur la page suivante]

WO 2004/088636 A1

KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), eurasien (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), européen (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Déclaration en vertu de la règle 4.17 :

— *relative à la qualité d'inventeur (règle 4.17.iv)) pour US seulement*

Publiée :

— *avec rapport de recherche internationale*

— *avant l'expiration du délai prévu pour la modification des revendications, sera republiée si des modifications sont reçues*

En ce qui concerne les codes à deux lettres et autres abréviations, se référer aux "Notes explicatives relatives aux codes et abréviations" figurant au début de chaque numéro ordinaire de la Gazette du PCT.

(57) Abrégé : Un système de reconnaissance de parole distribuée comporte au moins un terminal utilisateur, qui comprend des moyens d'obtention d'un signal audio à reconnaître, des moyens de calcul de paramètres et des moyens de contrôle pour sélectionner un signal à émettre, et un serveur qui comprend des moyens de réception du signal, des moyens de calcul de paramètres, des moyens de reconnaissance, et des moyens de contrôle pour commander les moyens de calcul et les moyens de reconnaissance en fonction du signal reçu.

SYSTEME DE RECONNAISSANCE DE PAROLE DISTRIBUEE

La présente invention est relative au domaine de la commande vocale d'applications, exercée sur des terminaux utilisateurs, grâce à la mise en oeuvre de moyens de reconnaissance de la parole. Les terminaux utilisateurs considérés sont tous les dispositifs dotés d'un moyen de capture de la parole, communément un microphone, possédant des capacités de traitement de ce son et reliés à un ou des serveurs par un canal de transmission. Il s'agit par exemple d'appareils de commande, de télécommande utilisés dans des applications domotiques, dans des automobiles (commande d'auto-radio ou d'autres fonctions du véhicule), dans des PC ou des postes téléphoniques. Le champ des applications concernées est essentiellement celui où l'utilisateur commande une action, demande une information ou veut interagir à distance en utilisant une commande vocale. L'utilisation de commandes vocales n'exclut pas l'existence dans le terminal utilisateur d'autres moyens d'action (système multi-modal), et le retour d'informations, d'états ou de réponses peut également se faire sous forme combinée visuelle, sonore, olfactive et tout autre moyen humainement perceptif.

De manière générale, les moyens pour la réalisation de la reconnaissance de parole comprennent des moyens d'obtention d'un signal audio, des moyens d'analyse acoustique qui extraient des paramètres de modélisation et enfin des moyens de reconnaissance qui comparent ces paramètres de modélisation calculés à des modèles, et proposent la forme mémorisée dans les modèles qui peut être associée au signal de la façon la plus probable. Optionnellement des moyens de détection d'activité vocale VAD (" Voice Activation Detection ") peuvent être utilisés. Ils assurent la détection des séquences correspondant à de la parole et devant être reconnues. Ils extraient du signal audio en entrée, en-dehors des périodes d'inactivité vocale, des segments de parole, qui seront ensuite traités par les moyens de calcul des paramètres de modélisation.

Plus particulièrement, l'invention porte sur les interactions entre les trois modes de reconnaissance de la parole dits embarqué, centralisé et distribué.

Dans un mode de reconnaissance de parole embarquée, l'ensemble des moyens pour effectuer la reconnaissance de parole se trouvent au niveau du terminal utilisateur. Les limitations de ce mode de reconnaissance sont donc liées notamment à la puissance des processeurs embarqués, et à la mémoire disponible pour stocker les modèles de reconnaissance de parole. En contrepartie, ce mode autorise un fonctionnement autonome, sans connexion à un serveur, et à ce titre est voué à un fort développement lié à la réduction du coût de la capacité de traitement.

Dans un mode de reconnaissance de la parole centralisée, toute la procédure de reconnaissance de parole et les modèles de reconnaissance se trouvent et s'exécutent sur une machine, appelée généralement serveur vocal, accessible par le terminal utilisateur. Le terminal transmet simplement au serveur un signal de parole. Cette méthode est utilisée notamment dans les applications offertes par les opérateurs de télécommunication. Un terminal basique peut ainsi accéder à des services évolués, activés à la voix. De nombreux types de reconnaissance de parole (robuste, flexible, très grand vocabulaire, vocabulaire dynamique, parole continue, mono ou multi locuteurs, plusieurs langues, etc) peuvent être implémentés dans un serveur de reconnaissance de parole. En effet, les machines centralisées ont des capacités de stockage de modèles, des tailles de mémoire de travail et des puissances de calcul importantes et croissantes.

Dans un mode de reconnaissance de parole distribuée, les moyens d'analyse acoustique sont embarqués dans le terminal utilisateur, les moyens de reconnaissance étant au niveau du serveur. Dans ce mode distribué, une fonction de débruitage associée aux moyens de calcul des paramètres de modélisation peut être avantageusement réalisée à la source. Seuls les paramètres de modélisation sont transmis, ce qui permet un gain substantiel en débit de transmission, particulièrement intéressant pour les applications multimodales. De plus, le signal à reconnaître peut être mieux protégé contre les erreurs de transmission. Optionnellement on peut aussi

embarquer la détection d'activité vocale (VAD) pour ne transmettre les paramètres de modélisation que durant les séquences de parole, ce qui a pour avantage de réduire de manière importante la durée de transmission active. La reconnaissance de parole distribuée permet en outre de véhiculer sur le même canal de transmission des signaux de parole et de données, notamment texte, images ou vidéos. Le réseau de transmission peut être par exemple de type IP, GPRS, WLAN ou Ethernet. Ce mode permet également de bénéficier de procédures de protection et de correction contre les pertes de paquets constituant le signal transmis à destination du serveur. Cependant il nécessite la disponibilité de canaux de transmission de données, avec un protocole strict de transmission.

L'invention propose un système de reconnaissance de parole comportant des terminaux utilisateurs et des serveurs combinant les différentes fonctions offertes par les modes de reconnaissance de parole embarquée, centralisée et distribuée, pour offrir le maximum d'efficacité, de confort et d'ergonomie aux utilisateurs de services multi modaux où la commande vocale est utilisée.

Le brevet US 6 487 534-B1 décrit un système de reconnaissance de parole distribuée comportant un terminal utilisateur disposant des moyens de détection d'activité vocale, de moyens de calcul des paramètres de modélisation et de moyens de reconnaissance. Ce système comprend en outre un serveur disposant également de moyens de reconnaissance. Le principe décrit est de réaliser au moins une première phase de reconnaissance au niveau du terminal utilisateur. Dans une deuxième phase optionnelle, les paramètres de modélisation calculés au niveau du terminal sont envoyés à destination du serveur, afin notamment de déterminer cette fois grâce aux moyens de reconnaissance du serveur, une forme mémorisée dans les modèles de celui-ci et associée au signal envoyé.

Le but visé par le système décrit dans le document cité est de diminuer la charge au niveau du serveur. Cependant il s'ensuit que le terminal doit réaliser le calcul des paramètres de modélisation en local avant de les transmettre éventuellement à destination du serveur. Or il y a des circonstances où, pour des raisons de gestion de charge ou pour des raisons

applicatives, il est préférable de réaliser ce calcul au niveau du serveur.

Il s'ensuit également que les canaux utilisés pour la transmission des paramètres de modélisation à reconnaître, dans un système selon le document cité ci-dessus, doivent être impérativement des canaux aptes à transmettre ce type de données. Or lorsque de tels canaux au protocole très strict ne sont pas forcément disponibles en permanence sur le réseau de transmission. C'est pourquoi il est intéressant de pouvoir utiliser des canaux classiques de transmission de signaux audio, pour ne pas retarder ou bloquer le processus de reconnaissance entamé au niveau du terminal.

Un but de la présente invention est de proposer un système distribué qui soit moins affecté par les limitations citées ci-dessus.

Ainsi suivant un premier aspect, l'invention propose un système de reconnaissance de parole distribuée, comportant au moins un terminal utilisateur et au moins un serveur aptes à communiquer entre eux par l'intermédiaire d'un réseau de télécommunications, dans lequel le terminal utilisateur comprend :

- des moyens d'obtention d'un signal audio à reconnaître ;
- des premiers moyens de calcul de paramètres de modélisation du signal audio; et
- des premiers moyens de contrôle pour sélectionner au moins un signal à émettre à destination du serveur parmi le signal audio à reconnaître et un signal indiquant les paramètres de modélisation calculés,

et dans lequel le serveur comprend :

- des moyens de réception du signal sélectionné en provenance du terminal utilisateur ;
- des seconds moyens de calcul de paramètres de modélisation d'un signal d'entrée ;
- des moyens de reconnaissance pour associer au moins une forme mémorisée à des paramètres d'entrée ; et
- des seconds moyens de contrôle pour commander les seconds moyens de calcul et les moyens de reconnaissance de façon à

- lorsque le signal sélectionné reçu par les moyens de réception est de type audio, activer les seconds moyens de calcul de paramètres en leur adressant le signal sélectionné en tant que signal d'entrée, et adresser les paramètres calculés par les seconds moyens de calcul aux moyens de reconnaissance en tant que paramètres d'entrée, et
- lorsque le signal sélectionné reçu par les moyens de réception indique des paramètres de modélisation, adresser lesdits paramètres indiqués aux moyens de reconnaissance en tant que paramètres d'entrée.

Ainsi le système selon l'invention permet de transmettre depuis le terminal utilisateur à destination du serveur soit le signal audio (compressé ou non), soit le signal délivré par les moyens de calcul des paramètres de modélisation du terminal. Le choix du signal transmis peut être défini soit par le type d'applications en cours, soit par l'état du réseau, soit suite à une coordination entre les moyens de contrôle respectifs du terminal et du serveur.

Un système selon l'invention donne la capacité au terminal utilisateur de réaliser, en fonction par exemple de paramètres d'entrée dont les moyens de contrôle disposent à un instant donné, le calcul des paramètres de modélisation au niveau du terminal ou au niveau du serveur. Ce calcul peut également être réalisé en parallèle au niveau du terminal et au niveau du serveur.

Un système selon l'invention permet d'effectuer la reconnaissance vocale depuis des terminaux de différents types coexistant au sein d'un même réseau, par exemple :

- des terminaux ne disposant d'aucun moyen de reconnaissance local (ou dont le moyen de reconnaissance local est inactif), auquel cas le signal audio est envoyé pour reconnaissance à destination du serveur ;

- des terminaux disposant de moyens de détection d'activité vocale sans moyens de calcul de paramètres de modélisation, ni moyens de reconnaissance (ou dont les moyens de calcul de paramètres et les moyens de reconnaissance sont inactifs), et transmettant au serveur pour

reconnaissance un signal audio d'origine ou un signal audio représentatif de segments de parole extraits du signal audio en-dehors des périodes d'inactivité vocale,

- et des serveurs disposant par exemple uniquement de moyens de reconnaissance, sans moyens de calcul de paramètres de modélisation.

Avantageusement, les moyens d'obtention du signal audio du terminal utilisateur peuvent comprendre en outre des moyens de détection d'activité vocale pour extraire du signal audio d'origine, en-dehors des périodes d'inactivité vocale, des segments de parole. Les moyens de contrôle du terminal sélectionnent alors au moins un signal à émettre à destination du serveur, parmi un signal audio représentatif des segments de parole et le signal indiquant les paramètres de modélisation calculés.

Avantageusement les moyens de contrôle du terminal sont adaptés pour sélectionner au moins un signal à émettre à destination du serveur parmi au moins le signal audio d'origine, le signal audio indiquant les segments de parole extraits du signal audio d'origine et le signal indiquant des paramètres de modélisation calculés. Au niveau du serveur, les moyens de contrôle sont adaptés pour commander les moyens de calcul et les moyens de reconnaissance de façon à, lorsque le signal sélectionné reçu par les moyens de réception est représentatif des segments de parole extraits par les moyens de détection d'activité vocale du terminal, activer les moyens de calcul de paramètres du serveur en leur adressant le signal sélectionné en tant que signal d'entrée, et adresser les paramètres calculés par ces moyens de calcul aux moyens de reconnaissance en tant que paramètres d'entrée.

Dans un mode de réalisation préféré, le serveur comporte en outre des moyens de détection d'activité vocale pour extraire d'un signal reçu de type audio, en-dehors des périodes d'inactivité vocale, des segments de parole. Dans ce cas, au niveau du serveur, les moyens de contrôle sont adaptés pour commander les moyens de calcul et les moyens de reconnaissance de façon à

• lorsque le signal sélectionné reçu par les moyens de réception est de type audio :

- si le signal reçu de type audio est représentatif de segments de parole après détection d'activité vocale, activer les seconds moyens de calcul de paramètres en leur adressant le signal sélectionné en tant que signal d'entrée, puis adresser les paramètres calculés par les seconds moyens de calcul de paramètres aux moyens de reconnaissance en tant que paramètres d'entrée ;
- sinon activer les moyens de détection d'activité vocale du serveur en leur adressant le signal sélectionné en tant que signal d'entrée, puis adresser les segments extraits par les moyens de détection d'activité vocale aux seconds moyens de calcul de paramètres en tant que paramètres d'entrée, puis adresser les paramètres calculés par les seconds moyens de calcul de paramètres aux moyens de reconnaissance en tant que paramètres d'entrée ;

◦ lorsque le signal sélectionné reçu par les moyens de réception indique des paramètres de modélisation, adresser lesdits paramètres indiqués aux moyens de reconnaissance en tant que paramètres d'entrée.

Avantageusement, le terminal utilisateur comprend en outre des moyens de reconnaissance pour associer au moins une forme mémorisée à des paramètres d'entrée.

Dans ce dernier cas, les moyens de contrôle du terminal peuvent être adaptés pour sélectionner un signal à émettre à destination du serveur en fonction du résultat fourni par les moyens de reconnaissance du terminal. Et le terminal utilisateur peut comporter en outre des moyens de stockage adaptés pour stocker un signal au niveau du terminal, pour pouvoir, au cas où le résultat de la reconnaissance locale au terminal n'est pas satisfaisante, envoyer le signal pour reconnaissance par le serveur.

Avantageusement, les moyens de contrôle du terminal peuvent être adaptés pour sélectionner un signal à émettre à destination du serveur indépendamment du résultat fourni par des premiers moyens de reconnaissance.

Il faut noter que les moyens de contrôle d'un terminal peuvent passer de l'un à l'autre des deux modes décrits dans les deux paragraphes ci-dessus, en fonction par exemple du contexte applicatif, ou de l'état du réseau.

De préférence, les moyens de contrôle du serveur coopèrent avec les moyens de contrôle du terminal. Le terminal peut ainsi éviter d'envoyer à destination du serveur par exemple un signal audio s'il y a déjà une charge importante au niveau des moyens de calcul de paramètres du serveur. Dans un mode possible de réalisation, les moyens de contrôle du serveur sont configurés pour coopérer avec les moyens du terminal pour adapter le type de signaux envoyés par le terminal en fonction des capacités respectives du réseau, du serveur et du terminal.

Les moyens de calcul et de reconnaissance du terminal peuvent être normalisés ou propriétaires.

Dans un mode de réalisation préféré, certains au moins parmi les moyens de reconnaissance et de calcul de paramètres au niveau du terminal lui ont été fournis par téléchargement, sous forme de code exécutable par le processeur du terminal, par exemple depuis le serveur.

Selon un deuxième aspect, l'invention propose un terminal utilisateur pour mettre en œuvre un système de reconnaissance de parole distribuée selon l'invention.

Selon un troisième aspect, l'invention propose un serveur pour mettre en œuvre un système de reconnaissance de parole distribuée selon l'invention.

D'autres caractéristiques et avantages de l'invention apparaîtront encore à la lecture de la description qui va suivre. Celle-ci est purement illustrative et doit être lue en regard des dessins annexés sur lesquels :

- la figure unique est un schéma représentant un système dans un mode de réalisation de la présente invention.

Le système représenté sur la figure unique comporte un serveur 1 et un terminal utilisateur 2, qui communiquent entre eux par l'intermédiaire d'un réseau (non représenté) disposant de canaux pour la transmission de signaux de voix et de canaux pour la transmission de signaux de données.

Le terminal 2 comporte un microphone 4, qui recueille la parole à reconnaître d'un utilisateur sous forme d'un signal audio. Le terminal 2 comporte également un module de calcul de paramètres de modélisation 6, qui réalise de façon connue en soi une analyse acoustique permettant d'extraire les paramètres pertinents du signal audio, et éventuellement pouvant avantageusement réaliser une fonction de débruitage. Le terminal 2 comprend un contrôleur 8, qui sélectionne un signal parmi le signal audio et un signal indicatif des paramètres calculés par le module de calcul de paramètres 6. Il comprend en outre une interface 10 pour l'émission sur le réseau du signal sélectionné, à destination du serveur,

Le serveur 1 comporte une interface réseau 12 pour recevoir les signaux qui lui sont adressés, un contrôleur 14 qui analyse le signal reçu et le dirige ensuite sélectivement vers un module de traitement parmi plusieurs modules 16,18,20. Le module 16 est un détecteur d'activité vocale, qui assure la détection des segments correspondant à de la parole et devant être reconnus. Le module 18 assure le calcul de paramètres de modélisation de façon semblable au module de calcul 6 du terminal. Toutefois, le modèle de calcul peut être différent. Le module 20 exécute un algorithme de reconnaissance de type connu, par exemple à base de modèles de Markov cachés avec un vocabulaire par exemple supérieur à 100 000 mots. Ce moteur de reconnaissance 20 compare les paramètres en entrée à des modèles de parole qui représentent des mots ou des phrases, et détermine la meilleure forme associée, compte tenu de modèles syntaxiques qui décrivent les enchaînements de mots attendus, de modèles lexicaux qui précisent les différentes prononciations des mots, et de modèles acoustiques représentatifs des sons prononcés. Ces modèles sont par exemple multilocuteurs, capables de reconnaître, avec une bonne fiabilité, de la parole, indépendamment du locuteur.

Le contrôleur 14 commande le module de VAD 16, le module de calcul de paramètres 18 et le moteur de reconnaissance 20 de façon à :

a/ lorsque le signal reçu par l'interface de réception 12 est de type audio et n'indique pas des segments de parole obtenus par détection d'activité vocale, activer le module VAD 16 en lui adressant le signal reçu en tant que

signal d'entrée, puis adresser les segments de parole extraits par le module VAD 16 au module de calcul de paramètres 18 en tant que paramètres d'entrée, puis adresser les paramètres calculés par ces moyens de calcul de paramètres 18 au moteur de reconnaissance 20 en tant que paramètres d'entrée ;

b/ lorsque le signal reçu par l'interface de réception 12 est de type audio et indique des segments de parole après détection d'activité vocale, activer le module de calcul de paramètres 18 en lui adressant le signal reçu en tant que signal d'entrée, puis adresser les paramètres calculés par ce module de calcul de paramètres 18 au moteur de reconnaissance 20 en tant que paramètres d'entrée ;

c/ lorsque le signal reçu par l'interface de réception 12 indique des paramètres de modélisation, adresser lesdits paramètres indiqués au moteur de reconnaissance 20 en tant que paramètres d'entrée.

Par exemple, dans le cas où l'utilisateur du terminal 1 utilise une application permettant de demander des informations sur la bourse et énonce : « cours de clôture des trois derniers jours de la valeur Lambda », le signal audio correspondant est capturé par le microphone 4. Dans le mode de réalisation du système selon l'invention, ce signal est ensuite, par défaut, traité par le module de calcul de paramètres 6, puis un signal indiquant les paramètres de modélisation calculés est envoyé vers le serveur 1.

Quand par exemple des problèmes de disponibilité de canaux de données ou du module de calcul 6 surgissent, c'est le signal audio en sortie du microphone 4 que le contrôleur 8 sélectionne alors pour le transmettre à destination du serveur 1.

Le contrôleur peut aussi être adapté pour envoyer systématiquement un signal indiquant les paramètres de modélisation.

Le serveur réceptionne le signal avec l'interface de réception 12, puis réalise, pour effectuer la reconnaissance de parole sur le signal reçu, le traitement indiqué en a/ ou b/ si le signal envoyé par le terminal 1 est de type audio ou le traitement indiqué en c/ si le signal envoyé par le terminal 1 indique des paramètres de modélisation.

Le serveur selon l'invention est également apte à effectuer de la reconnaissance de parole sur un signal transmis par un terminal ne disposant pas de moyens de calcul de paramètres de modélisation, ni de moyens de reconnaissance et disposant éventuellement de moyens de détection d'activité vocale.

Avantageusement, dans un mode de réalisation de l'invention, le système peut comporter en outre un terminal utilisateur 22, qui comporte un microphone 24 similaire à celui du terminal 2, un module 26 de détection d'activité vocale. La fonction du module 26 est semblable à celle du module de détection d'activité vocale 16 du serveur 1, Toutefois le modèle de détection peut être différent. Le terminal 22 comporte un module de calcul de paramètres de modélisation 28, un moteur de reconnaissance 30 et un contrôleur 32. Il comprend une interface 10 pour l'émission sur le réseau, à destination du serveur, du signal sélectionné par le contrôleur 32.

Le moteur de reconnaissance 30 du terminal peut par exemple traiter un vocabulaire de moins de 10 mots. Il peut fonctionner en mode monolocuteur, et nécessiter une phase d'apprentissage préalable à partir de la voix de l'utilisateur.

La reconnaissance de parole peut s'effectuer de différentes façons :

- exclusivement au niveau du terminal, ou
- ou exclusivement au niveau du serveur, ou
- partiellement ou totalement au niveau du terminal et également, de manière alternative ou simultanée, partiellement ou totalement au niveau du serveur.

Quand un choix doit être effectué sur la forme finalement retenue entre une forme associée fournie par le module de reconnaissance du serveur et une forme associée fournie par ceux du terminal, il peut s'effectuer sur la base de différents critères, qui peuvent varier d'un terminal à l'autre, mais aussi d'une application à l'autre ou d'un contexte donné à un autre. Ces critères peuvent donner par exemple priorité à la reconnaissance effectuée au niveau du terminal, ou à la forme associée présentant le plus fort taux de probabilité, ou encore à la forme déterminée le plus rapidement.

La façon dont s'effectue cette reconnaissance peut être figée au niveau du terminal dans un mode donné. Ou elle peut varier en fonction notamment de critères liés à l'application concernée, à des problématiques de charge des différents moyens au niveau du terminal et du serveur, ou encore à des problématiques de disponibilité de canaux de transmission voix ou données. Les contrôleurs 32 et 14 situés respectivement au niveau du terminal et du serveur traduisent la façon dont doit s'effectuer la reconnaissance.

Le contrôleur 32 du terminal est adapté pour sélectionner un signal parmi le signal audio d'origine en sortie du microphone 24, un signal audio représentatif des segments de parole extraits par le module VAD 26 et un signal indiquant des paramètres de modélisation 28. Suivant les cas, le traitement au niveau du terminal se poursuivra ou non au-delà de l'étape de traitement du terminal délivrant le signal à émettre.

Par exemple, considérons un mode de réalisation dans lequel le module VAD 26 du terminal est conçu par exemple pour détecter rapidement des mots de commandes et le module VAD 16 du serveur peut être plus lent, mais est conçu pour détecter des phrases entières. Une application, dans laquelle le terminal 22 effectue une reconnaissance en local et de façon simultanée fait effectuer une reconnaissance par le serveur à partir du signal audio transmis, permet notamment de cumuler les avantages de chaque module de détection vocale.

Considérons à présent une application dans laquelle la reconnaissance est effectuée exclusivement en local (terminal) ou exclusivement distante (serveur centralisé), sur la base de mots-clés permettant la commutation :

La reconnaissance en cours est d'abord locale : l'utilisateur énonce : « appelle Antoine », Antoine figurant dans le répertoire local. Puis il énonce « messagerie », mot-clé qui est reconnu en local et qui fait basculer en reconnaissance par le serveur. La reconnaissance est maintenant distante. Il énonce « rechercher le message de Josiane ». Lorsque ledit message a été écouté, il énonce « terminé », mot-clé qui fait à nouveau basculer l'application en reconnaissance locale.

Le signal transmis au serveur pour y effectuer la reconnaissance était de type signal audio. Dans un autre mode de réalisation, il pourrait indiquer les paramètres de modélisation calculés dans le terminal.

Considérons maintenant une application dans laquelle la reconnaissance au niveau du terminal et celle au niveau du serveur sont alternées. La reconnaissance est d'abord effectuée au niveau du terminal 22 et le signal après détection vocale est stocké. Si la réponse est consistante, c'est-à-dire s'il n'y a pas de rejet du module de reconnaissance 30 et si le signal reconnu est valide du point de vue applicatif, l'applicatif local au terminal passe à la phase applicative suivante. Dans le cas contraire, le signal stocké est envoyée au serveur pour effectuer la reconnaissance sur un signal indiquant des segments de parole après détection d'activité vocale sur le signal audio (dans un autre mode de réalisation, ce sont les paramètres de modélisation qui pourraient être stockés)

Ainsi l'utilisateur énonce « appelle Antoine » ; l'ensemble du traitement au niveau du terminal 22 s'effectue avec stockage du signal. Le signal est reconnu avec succès en local. Il énonce alors « rechercher le message de Josiane » ; la reconnaissance au niveau du terminal échoue ; le signal stocké est alors transmis au serveur. Le signal est bien reconnu et le message demandé est joué.

Dans une autre application, la reconnaissance se fait simultanément au niveau du terminal et également, et ce indépendamment du résultat de la reconnaissance locale, au niveau du serveur. L'utilisateur énonce « appelle Antoine ». La reconnaissance se déroule aux deux niveaux. Comme le traitement local interprète la commande, le résultat distant n'est pas considéré. Puis l'utilisateur énonce « rechercher le message de Josiane » qui génère un échec en local, et qui est bien reconnu au niveau du serveur.

Dans un mode de réalisation, le moteur de reconnaissance 30 du terminal 22 est un programme exécutable téléchargé depuis le serveur par des moyens classiques de transfert de données.

Avantageusement, pour une application donnée du terminal 22, des modèles de reconnaissance du terminal peuvent être téléchargés ou mis à jour au cours d'une session applicative connectée au réseau.

D'autres ressources logicielles utiles à la reconnaissance de parole peuvent aussi être téléchargées depuis le serveur 1, comme le module 6,28 de calcul de paramètres de modélisation ou le détecteur d'activité vocale 26.

D'autres exemples pourraient être décrits, mettant en œuvre par exemple des applications liées aux voitures, à l'électroménager, multimédia.

Comme présenté dans les exemples de réalisation ci-dessus décrits, un système selon l'invention permet d'utiliser de façon optimisée les différentes ressources nécessaires au traitement de la reconnaissance de la parole et présentes au niveau du terminal et du serveur.

REVENDICATIONS

1. Système de reconnaissance de parole distribuée, comportant au moins un terminal utilisateur et au moins un serveur aptes à communiquer entre eux par l'intermédiaire d'un réseau de télécommunications, dans lequel le terminal utilisateur comprend :

- des moyens d'obtention d'un signal audio à reconnaître ;
- des premiers moyens de calcul de paramètres de modélisation du signal audio; et
- des premiers moyens de contrôle pour sélectionner au moins un signal à émettre à destination du serveur parmi le signal audio à reconnaître et un signal indiquant les paramètres de modélisation calculés,

et dans lequel le serveur comprend :

- des moyens de réception du signal sélectionné en provenance du terminal utilisateur ;
- des seconds moyens de calcul de paramètres de paramètres de modélisation d'un signal d'entrée ;
- des moyens de reconnaissance pour associer au moins une forme mémorisée à des paramètres d'entrée ; et
- des seconds moyens de contrôle pour commander les seconds moyens de calcul et les moyens de reconnaissance de façon à

• lorsque le signal sélectionné reçu par les moyens de réception est de type audio, activer les seconds moyens de calcul de paramètres en leur adressant le signal sélectionné en tant que signal d'entrée, et adresser les paramètres calculés par les seconds moyens de calcul aux moyens de reconnaissance en tant que paramètres d'entrée, et

• lorsque le signal sélectionné reçu par les moyens de réception indique des paramètres de modélisation, adresser lesdits paramètres indiqués aux moyens de reconnaissance en tant que paramètres d'entrée.

2. Système selon la revendication 1, dans lequel les moyens d'obtention du signal audio à reconnaître comprennent des moyens de détection d'activité vocale pour produire le signal à reconnaître sous forme d'extraits d'un signal audio d'origine, en-dehors de segment de parole de périodes d'inactivité vocale.

3. Système selon la revendication 2, dans lequel les premiers moyens de contrôle sont adaptés pour sélectionner le signal à émettre à destination du serveur parmi au moins le signal audio d'origine, le signal audio à reconnaître sous forme des segments extraits par les moyens de détection d'activité vocale et le signal indiquant des paramètres de modélisation calculés par les premiers moyens de calcul de paramètres.

4. Système selon l'une quelconque des revendications précédentes, dans lequel :

- le serveur comporte en outre des moyens de détection d'activité vocale pour extraire d'un signal de type audio en-dehors de périodes d'inactivité vocale des segments de parole ; et
- les seconds moyens de contrôle sont adaptés pour commander les seconds moyens de calcul et les moyens de reconnaissance lorsque le signal sélectionné reçu par les moyens de réception est de type audio de façon à

si le signal de type audio est représentatif de segments de parole après détection d'activité vocale, activer les seconds moyens de calcul de paramètres en leur adressant le signal sélectionné en tant que signal d'entrée, puis adresser les paramètres calculés par les seconds moyens de calcul de paramètres aux moyens de reconnaissance en tant que paramètres d'entrée ;

sinon activer les moyens de détection d'activité vocale du serveur en leur adressant le signal reçu en tant que signal d'entrée, puis adresser les segments extraits par les seconds moyens de détection d'activité vocale aux seconds moyens de calcul de paramètres en tant que signal d'entrée, puis adresser les paramètres calculés par les seconds moyens de calcul de paramètres aux moyens de reconnaissance en tant que paramètres d'entrée.

5. Système selon les revendications 1 à 4, dans lequel le terminal utilisateur comprend en outre des moyens de reconnaissance pour associer au moins une forme mémorisée aux paramètres de modélisation calculés par les premiers moyens de calcul.

6. Système selon la revendication 5, dans lequel les premiers moyens de contrôle sont adaptés pour sélectionner le signal à émettre à destination du serveur en fonction du résultat fourni par les moyens de reconnaissance du terminal.

7. Système selon l'une des revendications 5 à 6, dans lequel le terminal utilisateur comporte en outre des moyens de stockage adaptés pour stocker le signal audio à reconnaître ou les paramètres de modélisation calculés par les premiers moyens de calcul de paramètres.

8. Système selon la revendication 5, dans lequel les premiers moyens de contrôle sont adaptés pour sélectionner un signal à émettre à destination du serveur indépendamment du résultat fourni par des moyens de reconnaissance du terminal.

9. Terminal utilisateur pour mettre en œuvre un système de reconnaissance de parole distribuée selon l'une des revendications 1 à 8, comportant :

- des moyens d'obtention d'un signal audio à reconnaître ;
- des moyens de calcul de paramètres de modélisation du signal audio ; et
- des premiers moyens de contrôle pour sélectionner au moins un signal à émettre à destination d'un serveur parmi le signal audio à reconnaître et un signal indiquant des paramètres de modélisation calculés.

10. Terminal utilisateur selon la revendication 9, dans lequel au moins une partie des moyens de calcul de paramètres est téléchargée depuis le serveur.

11. Terminal selon la revendication 9 ou 10 comprenant en outre des moyens de reconnaissance pour associer au moins une forme mémorisée aux paramètres de modélisation.

12. Terminal utilisateur selon la revendication 11, dans lequel au moins une partie des moyens de reconnaissance est téléchargée depuis le serveur.

13. Serveur pour mettre en œuvre un système de reconnaissance de parole distribuée selon l'une des revendications 1 à 8 comprenant :

- des moyens de réception, en provenance d'un terminal utilisateur, d'un signal sélectionné audit terminal ;
- des moyens de calcul de paramètres de modélisation d'un signal d'entrée ;
- des moyens de reconnaissance pour associer au moins une forme mémorisée à des paramètres d'entrée ; et
- des moyens de contrôle pour commander les seconds moyens de calcul et les moyens de reconnaissance de façon à

• lorsque le signal sélectionné reçu par les moyens de réception est de type audio, activer les moyens de calcul de paramètres en leur adressant le signal sélectionné en tant que signal d'entrée, et adresser les paramètres calculés par les moyens de calcul aux moyens de reconnaissance en tant que paramètres d'entrée, et

• lorsque le signal sélectionné reçu par les moyens de réception indique des paramètres de modélisation, adresser lesdits paramètres indiqués aux moyens de reconnaissance en tant que paramètres d'entrée.

14. Serveur selon la revendication 13 comprenant des moyens pour télécharger des ressources logicielles de reconnaissance vocale par l'intermédiaire du réseau de télécommunications à destination d'un terminal au

moins une partie des premiers moyens de calcul de paramètres ou des moyens de reconnaissance du terminal.

15. Serveur selon la revendication 14 comprenant des moyens pour télécharger des ressources logicielles de reconnaissance vocale par l'intermédiaire du réseau de télécommunications à destination d'un terminal.

16. Serveur selon la revendication 15, dans lequel lesdites ressources comprennent au moins un module parmi : un module de VAD, un module de calcul de paramètres de modélisation d'un signal audio et un module de reconnaissance pour associer au moins une forme mémorisée à des paramètres de modélisation.

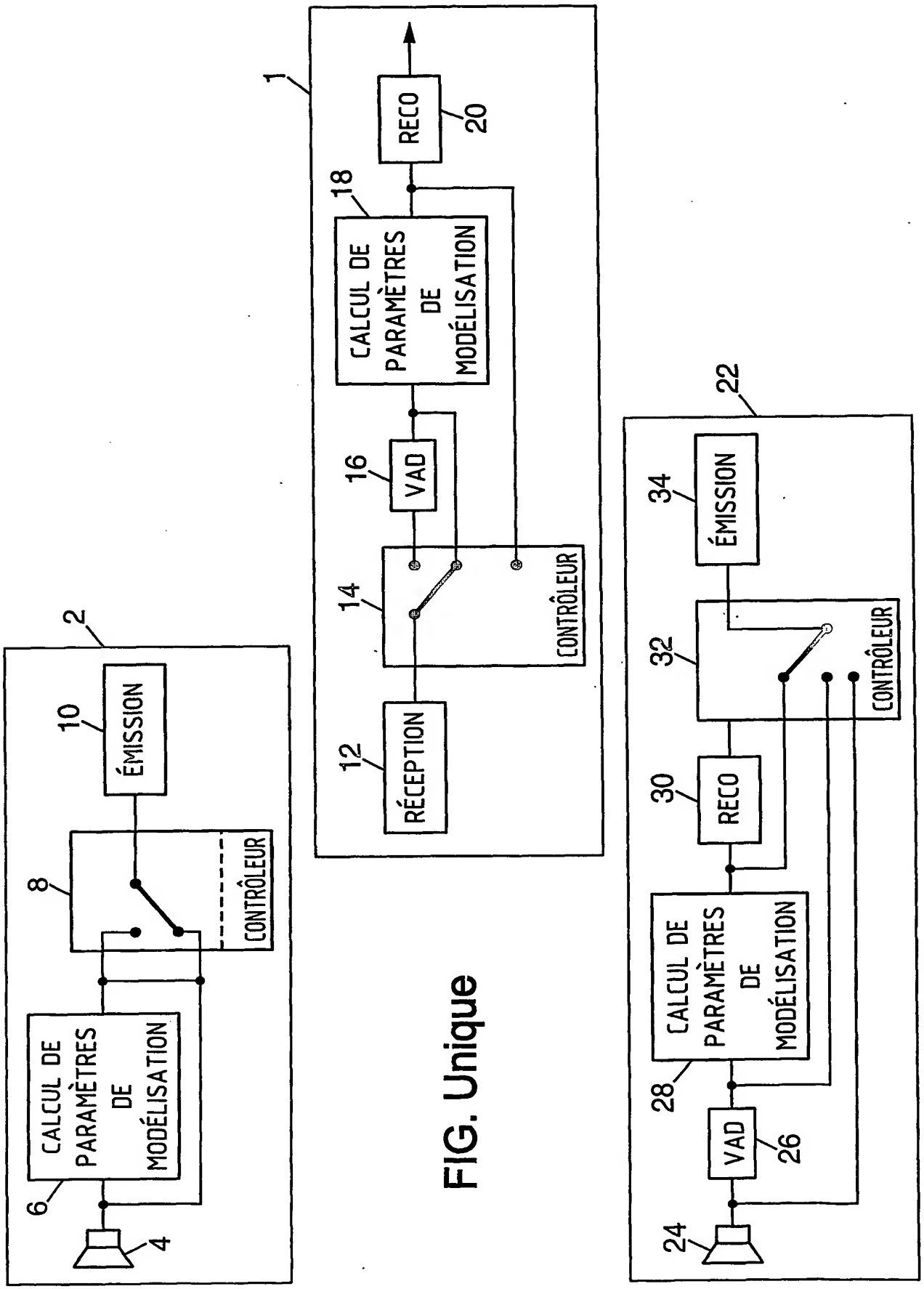

1
RECO
20
18
CALCUL DE PARAMÈTRES DE MODÉLISATION
16
VAD
14
CONTRÔLEUR
12
RÉCEPTION
2
10
ÉMISSION
8
CONTRÔLEUR
CALCUL DE PARAMÈTRES DE MODÉLISATION
6
4
22
34
ÉMISSION
32
CONTRÔLEUR
30
RECO
CALCUL DE PARAMÈTRES DE MODÉLISATION
28
VAD
26
24

FIG. Unique

INTERNATIONAL SEARCH REPORT

International Application No
PCT/FR2004/000546

A. CLASSIFICATION OF SUBJECT MATTER
IPC 7 G10L15/28

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 7 G10L

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)
EPO-Internal, WPI Data, INSPEC

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	EP 1 006 509 A (LUCENT TECHNOLOGIES INC) 7 June 2000 (2000-06-07) the whole document	1-16
Y	US 6 487 534 B1 (THELEN ET AL) 26 November 2002 (2002-11-26) cited in the application the whole document	1-9,13
Y	US 6 308 158 B1 (KUHNEN ET AL) 23 October 2001 (2001-10-23) abstract; figure 3	10-12, 14-16
A	US 6 122 613 A (BAKER JAMES K) 19 September 2000 (2000-09-19) the whole document	1-9,11, 13

-/--

[X] Further documents are listed in the continuation of box C.

[X] Patent family members are listed in annex.

° Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
28 July 2004	09/08/2004

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL – 2280 HV Rijswijk Tel. (+31–70) 340–2040, Tx. 31 651 epo nl, Fax: (+31–70) 340–3016	Quélavoine, R

INTERNATIONAL SEARCH REPORT

International Application No
PCT/FR2004/000546

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	WO 01/95312 A (NOKIA MOBILE PHONES LTD ;NOKIA INC (US)) 13 December 2001 (2001-12-13) abstract	1,5-9,13
A	WEIQI ZHANG ET AL: "The study on distributed speech recognition system" 2000 IEEE INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH, AND SIGNAL PROCESSING. PROCEEDINGS. (ICASSP). ISTANBUL, TURKEY, JUNE 5-9, 2000, IEEE INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH, AND SIGNAL PROCESSING (ICASSP), NEW YORK, NY: IEEE, US, vol. 3 OF 6, 5 June 2000 (2000-06-05), pages 1431-1434, XP002233412 ISBN: 0-7803-6294-2 abstract	1-16

INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No
PCT/FR2004/000546

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
EP 1006509 A	07-06-2000	US 6336090 B1	01-01-2002
		AU 5958599 A	01-06-2000
		CA 2287025 A1	30-05-2000
		DE 69911723 D1	06-11-2003
		EP 1006509 A1	07-06-2000
		JP 2000187496 A	04-07-2000
US 6487534 B1	26-11-2002	AU 3165000 A	16-10-2000
		CN 1351745 T	29-05-2002
		WO 0058946 A1	05-10-2000
		WO 0058942 A2	05-10-2000
		EP 1181684 A1	27-02-2002
		EP 1088299 A2	04-04-2001
		JP 2002540477 T	26-11-2002
		JP 2002540479 T	26-11-2002
US 6308158 B1	23-10-2001	AU 5904300 A	31-01-2001
		CA 2378215 A1	04-01-2001
		EP 1208562 A1	29-05-2002
		WO 0101391 A1	04-01-2001
US 6122613 A	19-09-2000	DE 69814589 D1	18-06-2003
		DE 69814589 T2	25-03-2004
		EP 0954848 A1	10-11-1999
		WO 9834217 A1	06-08-1998
WO 0195312 A	13-12-2001	AU 5059101 A	17-12-2001
		EP 1290678 A1	12-03-2003
		WO 0195312 A1	13-12-2001

RAPPORT DE RECHERCHE INTERNATIONALE

De[redacted]de internationale No
PCT/FR2004/000546

A. CLASSEMENT DE L'OBJET DE LA DEMANDE
CIB 7 G10L15/28

Selon la classification internationale des brevets (CIB) ou à la fois selon la classification nationale et la CIB

B. DOMAINES SUR LESQUELS LA RECHERCHE A PORTE

Documentation minimale consultée (système de classification suivi des symboles de classement)
CIB 7 G10L

Documentation consultée autre que la documentation minimale dans la mesure où ces documents relèvent des domaines sur lesquels a porté la recherche

Base de données électronique consultée au cours de la recherche internationale (nom de la base de données, et si réalisable, termes de recherche utilisés)
EPO-Internal, WPI Data, INSPEC

C. DOCUMENTS CONSIDERES COMME PERTINENTS

Catégorie °	Identification des documents cités, avec, le cas échéant, l'indication des passages pertinents	no. des revendications visées
Y	EP 1 006 509 A (LUCENT TECHNOLOGIES INC) 7 juin 2000 (2000-06-07) le document en entier	1-16
Y	US 6 487 534 B1 (THELEN ET AL) 26 novembre 2002 (2002-11-26) cité dans la demande le document en entier	1-9,13
Y	US 6 308 158 B1 (KUHNEN ET AL) 23 octobre 2001 (2001-10-23) abrégé; figure 3	10-12, 14-16
A	US 6 122 613 A (BAKER JAMES K) 19 septembre 2000 (2000-09-19) le document en entier	1-9,11, 13

-/--

[X] Voir la suite du cadre C pour la fin de la liste des documents

[X] Les documents de familles de brevets sont indiqués en annexe

° Catégories spéciales de documents cités:

"A" document définissant l'état général de la technique, non considéré comme particulièrement pertinent
"E" document antérieur, mais publié à la date de dépôt international ou après cette date
"L" document pouvant jeter un doute sur une revendication de priorité ou cité pour déterminer la date de publication d'une autre citation ou pour une raison spéciale (telle qu'indiquée)
"O" document se référant à une divulgation orale, à un usage, à une exposition ou tous autres moyens
"P" document publié avant la date de dépôt international, mais postérieurement à la date de priorité revendiquée

"T" document ultérieur publié après la date de dépôt international ou la date de priorité et n'appartenenant pas à l'état de la technique pertinent, mais cité pour comprendre le principe ou la théorie constituant la base de l'invention
"X" document particulièrement pertinent; l'inven tion revendiquée ne peut être considérée comme nouvelle ou comme impliquant une activité inventive par rapport au document considéré isolément
"Y" document particulièrement pertinent; l'inven tion revendiquée ne peut être considérée comme impliquant une activité inventive lorsque le document est associé à un ou plusieurs autres documents de même nature, cette combinaison étant évidente pour une personne du métier
"&" document qui fait partie de la même famille de brevets

Date à laquelle la recherche internationale a été effectivement achevée	Date d'expédition du présent rapport de recherche internationale
28 juillet 2004	09/08/2004

Nom et adresse postale de l'administration chargée de la recherche internationale	Fonctionnaire autorisé
Office Européen des Brevets, P.B. 5818 Patentlaan 2 NL – 2280 HV Rijswijk Tel. (+31–70) 340–2040, Tx. 31 651 epo nl, Fax: (+31–70) 340–3016	Quélavoine, R

RAPPORT DE RECHERCHE INTERNATIONALE

De[illegible]de Internationale No
PCT/FR2004/000546

C.(suite) DOCUMENTS CONSIDERES COMME PERTINENTS

Catégorie °	Identification des documents cités, avec, le cas échéant, l'indication des passages pertinents	no. des revendications visées
A	WO 01/95312 A (NOKIA MOBILE PHONES LTD ;NOKIA INC (US)) 13 décembre 2001 (2001-12-13) abrégé	1,5-9,13
A	WEIQI ZHANG ET AL: "The study on distributed speech recognition system" 2000 IEEE INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH, AND SIGNAL PROCESSING. PROCEEDINGS. (ICASSP). ISTANBUL, TURKEY, JUNE 5-9, 2000, IEEE INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH, AND SIGNAL PROCESSING (ICASSP), NEW YORK, NY: IEEE, US, vol. 3 OF 6, 5 juin 2000 (2000-06-05), pages 1431-1434, XP002233412 ISBN: 0-7803-6294-2 abrégé	1-16

RAPPORT DE RECHERCHE INTERNATIONALE

Renseignements rela... ...embres de familles de brevets

De... de Internationale No
PCT/FR2004/000546

Document brevet cité au rapport de recherche		Date de publication	Membre(s) de la famille de brevet(s)		Date de publication
EP 1006509	A	07-06-2000	US	6336090 B1	01-01-2002
			AU	5958599 A	01-06-2000
			CA	2287025 A1	30-05-2000
			DE	69911723 D1	06-11-2003
			EP	1006509 A1	07-06-2000
			JP	2000187496 A	04-07-2000
US 6487534	B1	26-11-2002	AU	3165000 A	16-10-2000
			CN	1351745 T	29-05-2002
			WO	0058946 A1	05-10-2000
			WO	0058942 A2	05-10-2000
			EP	1181684 A1	27-02-2002
			EP	1088299 A2	04-04-2001
			JP	2002540477 T	26-11-2002
			JP	2002540479 T	26-11-2002
US 6308158	B1	23-10-2001	AU	5904300 A	31-01-2001
			CA	2378215 A1	04-01-2001
			EP	1208562 A1	29-05-2002
			WO	0101391 A1	04-01-2001
US 6122613	A	19-09-2000	DE	69814589 D1	18-06-2003
			DE	69814589 T2	25-03-2004
			EP	0954848 A1	10-11-1999
			WO	9834217 A1	06-08-1998
WO 0195312	A	13-12-2001	AU	5059101 A	17-12-2001
			EP	1290678 A1	12-03-2003
			WO	0195312 A1	13-12-2001